



# An Enrichment of CRISPR and Other Defense-Related Features in Marine Sponge-Associated Microbial Metagenomes

Hannes Horn<sup>1,2</sup>, Beate M. Slaby<sup>1,2</sup>, Martin T. Jahn<sup>1,2</sup>, Kristina Bayer<sup>1</sup>, Lucas Moitinho-Silva<sup>3</sup>, Frank Förster<sup>4,5</sup>, Usama R. Abdelmohsen<sup>2,6</sup> and Ute Hentschel<sup>1,7\*</sup>

<sup>1</sup> RD3 Marine Microbiology, GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany, <sup>2</sup> Department of Botany II, Julius-von-Sachs Institute for Biological Sciences, University of Würzburg, Würzburg, Germany, <sup>3</sup> School of Biological, Earth and Environmental Sciences, Centre for Marine Bio-Innovation and School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia, <sup>4</sup> Department of Bioinformatics, University of Würzburg, Würzburg, Germany, <sup>5</sup> Center for Computational and Theoretical Biology, University of Würzburg, Würzburg, Germany, <sup>6</sup> Department of Pharmacognosy, Faculty of Pharmacy, Minia University, Minia, Egypt, <sup>7</sup> Christian-Albrechts-Universität zu Kiel, Kiel, Germany

## OPEN ACCESS

### Edited by:

Joerg Graf,  
University of Connecticut, USA

### Reviewed by:

Mark J. Mandel,  
Northwestern University, USA  
Kimberly B. Ritchie,  
Mote Marine Laboratory, USA

### \*Correspondence:

Ute Hentschel  
uhentschel@geomar.de

### Specialty section:

This article was submitted to  
Microbial Symbioses,  
a section of the journal  
Frontiers in Microbiology

Received: 28 July 2016

Accepted: 19 October 2016

Published: 08 November 2016

### Citation:

Horn H, Slaby BM, Jahn MT, Bayer K, Moitinho-Silva L, Förster F, Abdelmohsen UR and Hentschel U (2016) An Enrichment of CRISPR and Other Defense-Related Features in Marine Sponge-Associated Microbial Metagenomes. *Front. Microbiol.* 7:1751. doi: 10.3389/fmicb.2016.01751

Many marine sponges are populated by dense and taxonomically diverse microbial consortia. We employed a metagenomics approach to unravel the differences in the functional gene repertoire among three Mediterranean sponge species, *Petrosia ficiformis*, *Sarcotragus foetidus*, *Aplysina aerophoba* and seawater. Different signatures were observed between sponge and seawater metagenomes with regard to microbial community composition, GC content, and estimated bacterial genome size. Our analysis showed further a pronounced repertoire for defense systems in sponge metagenomes. Specifically, clustered regularly interspaced short palindromic repeats, restriction modification, DNA phosphorothioation and phage growth limitation systems were enriched in sponge metagenomes. These data suggest that defense is an important functional trait for an existence within sponges that requires mechanisms to defend against foreign DNA from microorganisms and viruses. This study contributes to an understanding of the evolutionary arms race between viruses/phages and bacterial genomes and it sheds light on the bacterial defenses that have evolved in the context of the sponge holobiont.

**Keywords:** metagenomes, defense, CRISPR, restriction modification, sponge microbiome, seawater

## INTRODUCTION

Marine sponges (Porifera) represent the oldest metazoan phylum with a fossil record dating back 580 million years in time (Li et al., 1998). Many sponges host dense and diverse communities of unicellular microorganisms within their tissues (Taylor et al., 2007; Hentschel et al., 2012; Thomas et al., 2016). Based on 16S rRNA gene amplicon sequencing, a recent study observed 1000s of symbiont lineages [operational taxonomic units (OTUs)] within sponges, which are dominated by *Proteobacteria* (mostly *Alpha*- and *Gammaproteobacteria*), *Acidobacteria*, *Actinobacteria*,

**Abbreviations:** CRISPR, clustered regularly interspaced short palindromic repeats; MTase, methyltransferase; nt, NCBI nucleotide database; nr, NCBI non-redundant protein database; REase, restriction endonuclease; RMS, restriction modification system.

*Chloroflexi*, *Cyanobacteria*, *Crenarchaeota*, as well as symbionts of several candidate phyla. Representatives of 41 different phyla were thus far recovered from sponges with representatives of 13 phyla being shared among all sponge hosts (Thomas et al., 2016). Sponges are ecologically important in benthic environments (Bell, 2008). The sponge-associated microorganisms carry out functions related to nutrient cycling including carbon, nitrogen, and possibly sulfur and vitamin metabolism (Taylor et al., 2007; Bayer et al., 2008; Hentschel et al., 2012) as well as to secondary metabolism and chemical defense (Wilson et al., 2014). As sessile filter feeders, sponges are capable of pumping seawater at rates up to 1000s of liters per kilogram of sponge per day (Vogel, 1977; Weisz et al., 2008). Small particles are retained from the incoming seawater and transferred into the mesohyl interior where they are digested by phagocytosis (Bell, 2008; Southwell et al., 2008; Maldonado et al., 2012). Sponges and their microbial consortia (hereafter referred to as the sponge holobiont) are thus continuously exposed to incoming microorganisms, that serve as a food source, but that may also be harmful (Webster, 2007; Wehrli et al., 2007). Despite considerable research effort and several published sponge genomes (Srivastava et al., 2010; Ryu et al., 2016), little is known as to how the sponge holobiont protects itself against potentially harmful microorganisms, whether eukaryotic, prokaryotic, or viral in nature.

One major line of prokaryotic defense is based on the self – non-self-discrimination principle, which recognizes and targets foreign DNA (Makarova et al., 2013). It comprises various systems, among them the clustered regularly interspaced short palindromic repeats (CRISPR). CRISPRs are based on conserved repeats and variable spacer sequences which are incorporated into the host genomes upon encounters from viruses or phages and plasmids and are thus able to memorize the attack (Horvath and Barrangou, 2010). Hence, it is described as the adaptive immune system of prokaryotes (Makarova et al., 2013). Structurally, CRISPRs are associated with *cas* genes, which are essential for their function and which are also used for the CRISPR classification (Makarova et al., 2011). Additional defense systems are the RMS and the DNA phosphothiolation (DND) system (Makarova et al., 2013). The RMS is nearly ubiquitous among bacteria (Vasu and Nagaraja, 2013). RMS can be classified into types I–IV depending on their subunits, recognition sites, cleavage positions, and substrate specificities (Roberts et al., 2003). Both, the RMS and DND systems, make use of labeling own DNA, either by methylation or by phosphorothioation, and recognize and destroy unmodified non-self DNA (Wang et al., 2007; Vasu and Nagaraja, 2013). The Phage growth limitation (Pgl) system is another line of defense that allows phage burst upon initial infection. In *Streptomyces coelicolor* A(3)2, Pgl was shown to target phage  $\Phi$ 31 and its relatives. Here, the DNA of the phage progeny was methylated, which resulted in activation and consequently, in prevention of phage growth through presumed methyl-specific restriction endonuclease activity (Abeldon, 2012; Hoskisson et al., 2015). The PglZ protein family is a central element of Pgl, however, the mechanisms of this complex system are poorly understood (Makarova et al., 2013). Another major line of defense is based on dormancy or programmed cell death

(Makarova et al., 2013). These can be separated into toxin–antitoxin (T–A) systems and abortive infection (ABI). In the T–A system, the protein toxin kills cells above a certain expression level. The antitoxin component then regulates and/or inactivates toxin expression and prevents killing of the cell. The ABI system is also based on cell death or dormancy and it is also based on two modules (Fineran et al., 2009). The ABI system activates cell death to prevent viral replication and thereby protects the bacterial population.

In the present study we aimed to characterize defense systems of marine sponge-associated microbial consortia. The microbial metagenomes of three Mediterranean sponges (*Petrosia ficiformis*, *Sarcotragus foetidus*, *Aplysina aerophoba*) and seawater were compared toward this goal. Besides insights into the microbial community composition and overall GC content, we present defense-related features that consist of the CRISPR system, restriction modification, phage growth inhibition, and genes related to DNA phosphothiolation. The results of the present study are consistent with the concept of “functional convergence” (Fan et al., 2012) that shows similar functional profiles in the microbiomes of different sponge species and that are distinct from those of seawater.

## MATERIALS AND METHODS

The sponges *P. ficiformis* (sample ID: 1Biotec2\_S07) and *S. foetidus* (sample ID: 1Biotec2\_S06) were collected on 25 May 2013, by SCUBA diving in Milos, Greece (N36.76759° E24.51422°), at 5–7 m depth. Sponge tissues (5 ml each) were washed with sterile-filtered seawater, passed through a 100  $\mu$ m Nitex cloth (Hartenstein, Germany) and transported to the laboratory in glycerol solution (15% v/v) at  $-20^{\circ}\text{C}$  until further processing. A total of 10 L seawater (sample ID: Biotec\_SW) was collected from the vicinity of the sponges. Within 2–3 h after collection, seawater was filtered consecutively through 100  $\mu$ m Nitex (Hartenstein), 5  $\mu$ m durapore (Merck-Millipore), and finally through 0.22  $\mu$ m durapore membrane filters, which were then frozen at  $-20^{\circ}\text{C}$ .

Sponge samples of *A. aerophoba* were collected in the Mediterranean Sea from a depth of 5 m (Piran, Slovenia), on 07 May 2013. Upon transport back to the laboratory, samples of pinacoderm and mesohyl were separated using a sterile scalpel blade. One scalpel blade was used per each sample to prevent cross-contamination between samples. Microbial cells were enriched from the different sponge tissues by differential centrifugation (Fieseler et al., 2006). Microbes from *P. ficiformis* and *S. foetidus* samples were prepared using the same protocol. Fractions of sponge-associated prokaryotes (SAPs) were frozen at  $-80^{\circ}\text{C}$  in 15% glycerin.

## DNA Extraction and Sequencing

Genomic DNA was extracted from the sponge SAP preparations of *P. ficiformis* and *S. foetidus* and the seawater filters using the FastDNA Spin Kit for Soil (MP Biomedicals, USA). The quantity of metagenomic DNA was determined by spectrophotometry using a NanoDrop 2000c reader (PEQLAB Biotechnologie

GmbH, Germany). The quality and size were analyzed by visual inspection on 0.8% agarose gels following electrophoresis.

DNA of *A. aerophoba* was extracted in triplicates for each pinacoderm and mesohyl using the FastDNA SPIN Kit for Soil (MP Biomedicals). In order to maximize DNA yield from bacteria with different cell properties, the cell lysis step varied for the three replicates of each tissue type: (i) bead beating, following the manufacturer's protocol, (ii) freeze-thaw cycling (three cycles of 20 min at  $-80^{\circ}\text{C}$  and 20 min at  $42^{\circ}\text{C}$ ), (iii) proteinase K digestion (bacterial pellet re-suspended in 567  $\mu\text{l}$  TE with SDS in a final concentration of 0.5% and proteinase K in 100 ng/ml final concentration) for 1 h at  $37^{\circ}\text{C}$ . After cell lysis, the manufacturer's protocol was followed for all six samples. Extracted metagenomic DNA from *A. aerophoba* samples was sequenced on an Illumina HiSeq2000 platform (150 bp paired-end reads) and quality filtered at the DOE Joint Genome Institute (Walnut Creek, CA, USA). Seawater, *P. ficiformis* and *S. foetidus* derived DNA was sequenced at GATC Biotech AG (Cologne, Germany) on an Illumina MiSeq Personal Sequencer (250 or 300 bp paired-end reads, respectively).

## Raw Data Processing and Assembly

The raw reads obtained for the samples of *P. ficiformis*, *S. foetidus*, and seawater were initially analyzed with FastQC 0.11.2<sup>1</sup> for adapters, overall quality, length and ambiguous bases. In a first step, the reads were trimmed using Trimmomatic 0.31 (PE -phred 33 LEADING:3 ILLUMINACLIP:2:30:10) (Bolger et al., 2014) and then merged using bbmerge<sup>2</sup>. All reads, merged and unmerged, were again subjected to Trimmomatic for further quality trimming and length filtering (SE -phred 33 SLIDINGWINDOW:4:25 MINLEN:150 AVGQUAL:30). The remaining reads were assembled with IDBA-UD 1.1.1 (-mink 10 -maxk 100) (Peng et al., 2012). Contigs with a length  $\leq 1000$  nt were discarded. The reads obtained for the *A. aerophoba* dataset were processed via the IMG/ER webserver (Markowitz et al., 2012). Quality filtered reads were normalized using bbnorm and assembled with SPAdes 3.5.0 (-only-assembler, -k 21,33,55,77,99,127, -sc) (Bankevich et al., 2012). Only contigs  $\geq 1000$  nt were used for further analysis. To remove eukaryotic contamination, all contigs, that were further analyzed, were subjected to blastn 2.2.28 (e-value  $10\text{e-}6$  -task blastn) (Altschul et al., 1990) and searched against the NCBI nucleotide database (nt, as of September 29, 2015). The blast hits were analyzed with Krona 2.6 (Ondov et al., 2011). All reads of eukaryotic origin were removed. Information about the metagenomics datasets is presented in **Tables 1** and **2**.

## Taxonomic Affiliation of Reads and Contigs

The processed reads were submitted to MG-RAST with enabled screening for human contamination and disabled dynamic trimming (Meyer et al., 2008). Contigs obtained from the metagenomic assemblies were assigned to taxonomy using blastx 2.2.28 (e-value  $10\text{e-}6$ ) and the NCBI non-redundant protein

database (nr). All hits were submitted to *blast2lca* (default parameters), a last common ancestor algorithm implemented in MEGAN5 (Huson et al., 2011).

## Comparison of GC Content and Average Genome Sizes

The GC content of all four metagenomes was calculated for all processed and filtered reads, using an in-house perl script. In addition, the average genome size per metagenome was computed with MicrobeCensus 1.0.7 (Nayfach and Pollard, 2015) using the same reads and their average calculated length (**Table 1**).

## Data Normalization

Processed reads were mapped to their respective assembly using bowtie2 2.2.4 (very-sensitive) (Langmead and Salzberg, 2012). The coverage for each position on a contig was calculated with samtools depth 1.2 (Li et al., 2009). With this data, the coverage of each contig was set as the mean coverage over each position. To account for the different sequencing depths, the number of mapped reads and assembly size, the coverage for each contig was divided by the total number of mapped basepairs and multiplied by  $10^6$  to obtain copy numbers per megabase (cpm).

## Functional Annotation

All contigs were subjected to Prodigal 2.6.0 (-p meta, -c, -g 11) (Hyatt et al., 2010) to predict open reading frames (ORFs). Clusters of Orthologous Groups (COGs) obtained from the Conserved Domains Database (CDD) (Marchler-Bauer et al., 2015) were annotated using rpsblast 2.2.28 (e-value  $10\text{e-}6$ ). Protein families (Pfam) and TIGRFAM were assigned with the InterProScan pipeline 5.17 (Jones et al., 2014) based on the best hit (e-value  $10\text{e-}6$ ).

## Characterization of CRISPR Arrays, Repeats, and Spacers

The presence of CRISPR arrays was analyzed with a multiple tool approach similar as proposed by Gogleva et al. (2014) using CRT, PILER-CR, and CRISPRFinder (Bland et al., 2007; Edgar, 2007; Grissa et al., 2007b). *Cas* genes were identified by subjecting ORFs of CRISPR-containing contigs to TIGRFAM and Pfam databases using InterProScan. Assignment of CRISPR-Cas types was accomplished according to Makarova et al. (2011) using the TIGRFAM and Pfam annotations. Contigs containing CRISPR arrays found with CRT and PILER-CR or included *cas* genes

**TABLE 1 |** Samples analyzed in this study.

	Seawater	<i>Petrosia ficiformis</i>	<i>Sarcotragus foetidus</i>	<i>Aplysina aerophoba</i>
Sample date	29.05.2013	29.05.2013	29.05.2013	07.05.2013
Location	Mediterranean Sea, Milos, Greece	Mediterranean Sea, Milos, Greece	Mediterranean Sea, Milos, Greece	Mediterranean Sea, Piran, Slovenia
Depth	5–7 m	5–7 m	5–7 m	5 m
Temperature	20°C	20°C	20°C	18°C

<sup>1</sup><http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

<sup>2</sup><https://sourceforge.net/projects/bbmap/>

**TABLE 2 | Statistics on the processing of the metagenomics samples from sequencing throughput to analysis.**

	Seawater	<i>Petrosia ficiformis</i>	<i>Sarcotragus foetidus</i>	<i>Aplysina aerophoba</i>
Sequencing platform	Illumina MiSeq (2 × 300 bp)	Illumina MiSeq (2 × 250 bp)	Illumina MiSeq (2 × 300 bp)	Illumina HiSeq (2 × 150 bp)
Sequenced reads (#)	40,505,000	41,383,600	32,672,426	945,906,728
Sequenced bp	12,151,500,000	10,345,900,000	9,801,727,800	283,772,018,400
Reads after QC (#)	18,273,997	29,213,518	17,525,606	–
Bp after QC	6,240,860,642	7,655,556,186	4,909,483,386	–
Assembly algorithm	IDBA-UD	IDBA-UD	IDBA-UD	SPAdes
Assembly size (bp)	216,407,276	226,772,563	190,159,175	489,999,481
Contigs > 1000 bp	116,626	82,740	41,164	110,609
N <sub>50</sub> contigs (bp)	1,853	3,381	9,706	8,958
Largest contig	58,177	342,148	369,775	1,056,271
Average %GC	41	63	63	58
Open reading frames	215,442	221,522	175,356	455,396
ORF with COG annotation	129,900	119,914	103,075	203,692
ORF with Pfam annotation	78,569	55,478	29,253	56,643
ORF with TIGRFAM annotation	28,017	17,214	10,675	17,267
Average genome size (bp)	1,347,075.38	3,034,048.52	3,744,502.76	5,165,191.54
Reads mapped to assembly	13,396,184	15,900,219	22,478,672	537,464,688
Bp mapped to assembly	3,642,606,507	3,539,701,337	5,378,691,619	80,619,703,200
Average coverage	16.83	15.61	28.29	164.53

were uploaded to CRISPRfinder and were validated as true hits. Of these, only confirmed CRISPR with at least two spacers were retained. Possible targets of spacers were identified by submitting their sequences to CRISPRtarget using the ACLAME (as of August, 2009), GenBank-Phage, GenBank-Plasmid and RefSeq-Viral databases (all as of September, 2015) (gap open -5, gap extend -2, nucleotide match +1, mismatch -1, e-value 0.1, word size 7) (Biswas et al., 2013). Direct repeat sequences were submitted to CRISPRdb (Grissa et al., 2007a) and blasted against the CRISPRfinder database (e-value 10e-2) and CRISPRmap (Lange et al., 2013) to examine their superclasses by sequence and structure and to determine if they were reported before. The origin of the CRISPR arrays was determined through their respective contigs as described in Section “Taxonomic Affiliation of Reads and Contigs.”

## Analysis of Restriction Modification Systems

Reference protein sequences of type I [restriction endonucleases (REase), methyltransferases (MTase), and specificity domains], type II (REases and MTases), and type III (REases and MTases) RMSs) were downloaded from REBASE (as of October 15, 2015) (Roberts et al., 2015). For each type of REases, MTases and specificity domains, a blast database was built. Predicted ORFs from all metagenomes were queried against the databases using blastp 2.2.28 (e-value 10e-6) and only hits with a coverage  $\geq 70\%$  were kept. A RMS was considered as being complete, if its restriction endonuclease and methyltransferase were at least four genes apart from each other (Oliveira et al., 2014). Finally, overlapping regions of REases and MTases (and specificity domains for type I) of the same type within four genes were combined to one cluster to avoid double counts.

## Deposition of Sequence Data

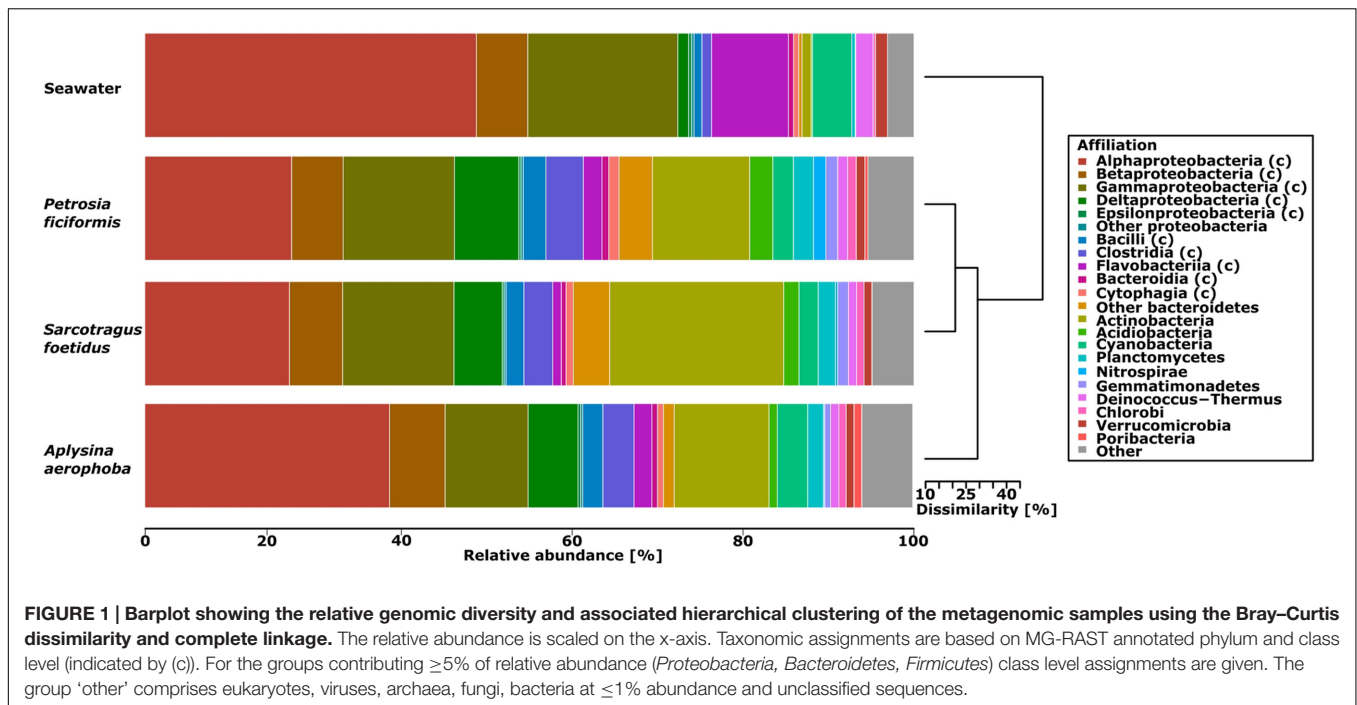
The sequencing projects were completed in 2013 and sequencing data was deposited in the Sequence Read Archive (SRA), metagenome assemblies as a Whole Metagenome Shotgun (WGS) projects in GenBank under the BioProject PRJNA318959 and the BioSample IDs SAMN04870510, SAMN04870527, SAMN04870528 and SAMN05860141 for *P. ficiformis* (SRA: SRP074318, WGS: LXNJ000000000), *S. foetidus* (SRA: SRP074318, WGS: LXNI000000000), seawater (SRA: SRP074318, WGS: LXNH000000000), and *A. aerophoba* (WGS: MKWU000000000). Raw sequencing data of *A. aerophoba* is available under the GOLD Study ID Gs0099546<sup>3</sup> with the GOLD Project IDs Gp005580–Gp005585 which can be downloaded via the JGI Genome Portal.

## RESULTS

### Sample Description

Three samples from the sponges *P. ficiformis*, *S. foetidus*, and *A. aerophoba* as well as one seawater sample from the Mediterranean Sea were investigated in this study for functional differences of their associated microbiomes (Table 1). Using Illumina MiSeq and HiSeq platforms, more than 1,064,000,000 high-quality sequences (~310 Gbp) were generated. The metagenomes had assembly sizes ranging from 190 to 489 Mbp. The predicted ORFs ranged from 175,356 in *S. foetidus* up to 455,396 in *A. aerophoba*. A total of 44.73–60.29% could be annotated via COGs (Table 2). In order to compare the generated data, the metagenomes were normalized based on their coverage which ranged from 15.61- to 164.53-fold.

<sup>3</sup><https://gold.jgi.doe.gov/biosamples?Study.GOLD%20Study%20ID=Gs0099546>



## Genomic Composition

Based on phylogenetic affiliations using the lowest common ancestor algorithm (LCA) in MG-RAST, over 95% of the reads of all four metagenome samples were assigned to 38 bacterial phyla. Proportions of archaea, eukaryotes, viruses and unclassified sequences were consistently low (each  $\leq 2.72\%$ ) in all metagenomes and were thus not further analyzed (Supplementary Figure S1). The genomic composition of the metagenomes was then analyzed on the phylum and class level. As indicated by Bray-Curtis dissimilarity, the metagenomes of *P. ficiformis* and *S. foetidus* were closest to each other (11.18% dissimilarity). Both showed dissimilarities of 17.01 and 19.46% to *A. aerophoba*. The seawater sample displayed dissimilarities of 41.04, 43.55, and 31.26% to *P. ficiformis*, *S. foetidus*, and *A. aerophoba*, respectively (Figure 1).

A limited number of sequences (0.15–0.21%) was not assigned to any known bacterial taxa. The *Proteobacteria*, *Firmicutes*, and *Bacteroidetes* were the most abundant phyla in all metagenomes. The *Actinobacteria* (12.3–22.63% vs. 1.18%) and the *Deltaproteobacteria* (6.28–7.24% vs. 1.38%) were more abundant in the sponge samples than in seawater. In contrast, the *Alphaproteobacteria* were less abundant in sponges than in seawater (18.80–31.81% vs. 43.11%), and so were the *Flavobacteria* (1.12–2.43% vs. 9.99%) and the *Cyanobacteria* (2.49–3.96% vs. 4.98%). Only minor differences between the sponge and seawater samples were found for the *Gammaproteobacteria* (10.77–14.48% vs. 19.51%), the *Clostridia* (3.75–4.86% vs. 1.25%) and other unclassified *Bacteroidetes* (1.38–4.74% vs. 0.43%) according to a principal component analysis (Supplementary Figure S3). Overall, the sponge metagenomes were taxonomically distinct from the seawater

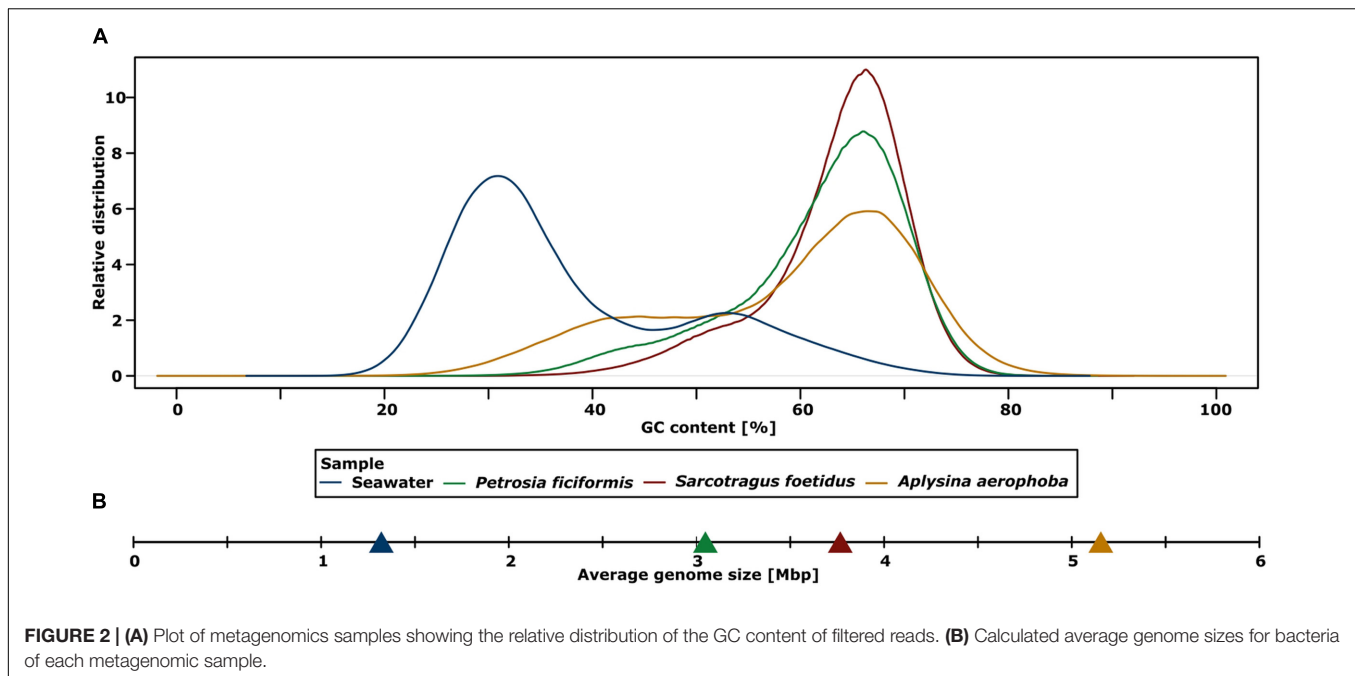
metagenome based on taxonomic read assignment using MG-RAST, Bray-Curtis dissimilarity, and principal component analysis.

## GC Footprint

Higher average GC contents were detected for the assembled metagenomes of sponges (58–63%) than for seawater (41%) (Figure 2A; Table 2). The highest GC content was detected for the metagenome sample of *S. foetidus*, followed by *P. ficiformis*, *A. aerophoba*, and seawater. Interestingly, a second smaller seawater peak around 50–55% overlapped with the lower GC tail ends of the sponge metagenomes. To test whether there is a correlation between high GC content and genome size, we calculated the average genome sizes for a bacterial cell within each metagenome. The calculated average genome sizes in the sponge sample were considerably higher than those of the seawater sample (Figure 2B; Table 2).

## General Functional Properties

Functional analysis was based on COG assignments. All hits were normalized to copy number per megabase based on their contig coverage. We identified 103,075–203,692 COG hits for the metagenomes which corresponds to 44.73–60.29% of annotated ORFs (Table 2). This number includes however the general function (G) and unknown function (S) categories (10.9–21.2 and 5.9–6.9%, respectively). The functional profiles of the sponge samples were more similar to each other than to seawater, as reflected by a Bray-Curtis dissimilarity of 10% between sponge and seawater metagenomes. Overall, many genes relate to the COG categories general function (G) or unknown function (S), and most of the COG categories were neither enriched for sponges nor the seawater metagenomes (“enriched” is defined as



**FIGURE 2 | (A)** Plot of metagenomics samples showing the relative distribution of the GC content of filtered reads. **(B)** Calculated average genome sizes for bacteria of each metagenomic sample.

>1.5-fold more copies per megabase) (Figure 3). The category nucleotide transport and metabolism (F) was exceptionally high in *S. foetidus* (33.05 cpm), whereas the category of cell cycle control, cell division, chromosome partitioning (D) was exceptionally low in *P. ficiformis* (0.53 cpm) when compared to the other sponge metagenomes. Only few differences were identified between seawater and sponge metagenomes based on COG level assignments. The sponge metagenomes showed a higher number of genes assigned to functions related to defense mechanisms (V) and the cytoskeleton (Z), suggesting that these are important functional traits for sponge symbionts. On the other hand, fewer reads were assigned to translation, ribosomal structure and biogenesis (O), cell motility (N), and chromatin structure and dynamics (B) in the sponge sample, marking them as relevant functional features for free living bacteria.

## Defense Mechanisms

### COG and Pfam-Annotated Defense Mechanisms

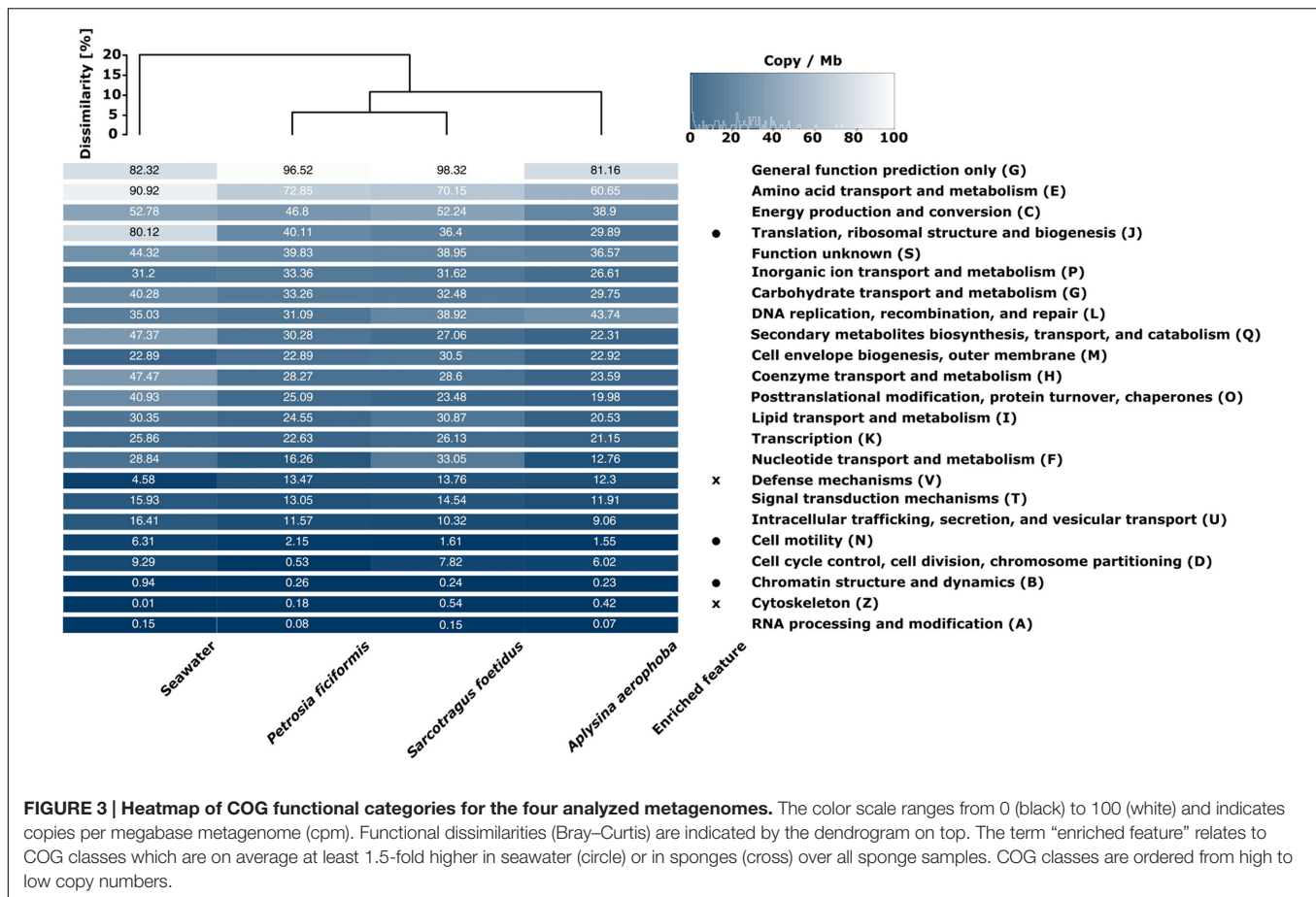
With respect to defense mechanisms, the sponge datasets were more similar to each other than to seawater according to Bray–Curtis dissimilarity measure (Figures 4A,B). All comparisons are based on copies per megabase (cpm). Features were defined as “enriched” when being >1.5-fold abundant in either the sponge or seawater metagenome. Transport and efflux systems for drugs were found in all samples, and with the exception of a Na<sup>+</sup>-efflux pump and ABC-type multidrug transporter, all related functions were enriched in the sponge samples over seawater. Furthermore, all annotations associated to CRISPR were enriched in the sponge microbiomes, as all (with the exception of one CRISPR-nuclease (COG3513) were absent from seawater. 11 features related to CRISPR in *S. foetidus* and one in *P. ficiformis* were missing from the sponge metagenomes, which were mostly related to the receptor activity-modifying proteins (RAMP) superfamily.

Interestingly, the *cas2*-gene (COG1343) in *S. foetidus* may be substituted by a *cas2*-homolog (COG3512), which showed highest cpm within this metagenome. With respect to RMSs, all genes except one encoding for one endonuclease (COG1403) were enriched in the sponge datasets. However, one endonuclease (COG1787) copy was absent in *S. foetidus* and six were absent from seawater. The overall cpm’s within the RMS were higher in sponge metagenomes than in the corresponding seawater metagenome. Classes A and C beta-lactamases (COG2367, COG1680) were further enriched in sponge metagenomes. On the other hand, seawater was enriched for the beta-lactamase class D (COG2602) and an inductive membrane protein (COG3725). A couple of genes related to resistance against colicin (COG4452), a growth inhibiting toxin, bacteriophages (COG4823) and the antibiotic vancomycin (COG2720) were enriched in sponge metagenomes, whereas a cephalosporin hydroxylase (COG3510) was more abundant in the seawater metagenome (Figure 4A).

The overall gene copy numbers for DNA phosphorothioation (DND) and phage growth limitation (Pgl) were higher in the sponge than in the seawater metagenome (Figure 4B). The DndG (PF08747) was absent from seawater. With respect to the Pgl system, two core genes (COG1002, COG4930) and three additional genes (PF08849, PF10923, COG3472) were missing from seawater. The overall Pgl gene copy number in sponges was ~50% higher in *A. aerophoba* and *S. foetidus* than in *P. ficiformis* (Figure 4B).

### Clustered Regularly Interspaced Short Palindromic Repeats

We analyzed CRISPR arrays and related components, i.e., direct repeat sequences separated by spacers and adjacent *cas* genes in the four metagenomes. The highest numbers of CRISPR array containing contigs were found by searching for *cas* genes.



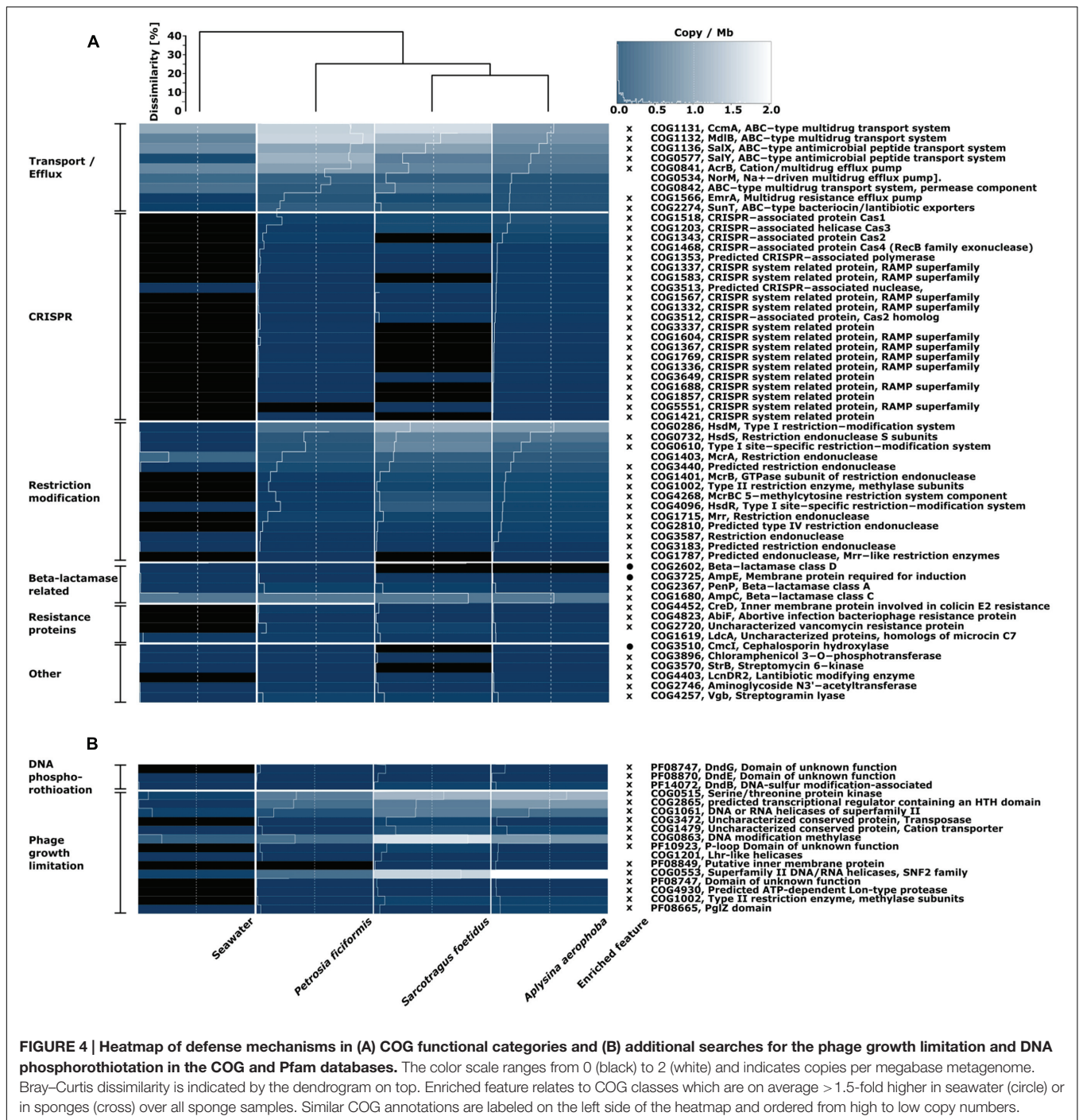
*A. aerophoba* showed the highest abundance of validated arrays, whereas none was identified in the seawater metagenome. The final number of identified CRISPR arrays was 77 (0.21 cpm), 47 (0.25 cpm), 283 (0.62 cpm), and 0 (0 cpm) for the metagenomes of *P. ficiformis*, *S. foetidus*, *A. aerophoba* and seawater, respectively (Table 3; Figure 5). On the domain level, taxonomy was assigned to 53 of 77 (68.83%) arrays in *P. ficiformis*, 40 of 47 (85.11%) arrays in *S. foetidus* and 240 of 283 (84.81%) arrays in *A. aerophoba*. Noteworthy, despite the differences in their geographic location and total number of identified arrays, a large overlap of taxonomic groups was found. The overall distribution of taxa containing CRISPR-contigs was similar in the three sponge datasets, with *Proteobacteria* as the most prevalent phylum followed by *Actinobacteria* and *Chloroflexi* in the sponges *P. ficiformis* and *S. foetidus* and *Firmicutes* in *A. aerophoba* (Table 3).

Different CRISPR-Cas types were categorized by their associated *cas* genes. In the metagenomic datasets, at least 26 (35.06%), 20 (42.55%), and 144 (46.78%) of the CRISPR arrays were adjacent to *cas* genes for *P. ficiformis*, *S. foetidus* and *A. aerophoba*, which indicates that these arrays might be complete (Table 3). The *cas* genes of all known CRISPR-Cas types were identified. CRISPR-Cas type I was the most abundant, with the subtypes I-E and I-C as the most prevalent for all metagenomes, followed by types II and III. Around 50% of *cas* genes could not

be annotated in more detail (type unknown, Table 3). According to the number of *cas* genes per megabase, *cas1*, *cas2* were most abundant, followed by *cas3*, *cas4* and *cas7* in all three sponge metagenomes. Smallest proportions were detected for *cas8* and *cas9* (Figure 5). Even though CRISPR arrays were not identified in the seawater metagenome, two *cas6*-genes were detected.

Spacers are the functional part of the CRISPR defense that recognizes foreign DNA fragments. In the *P. ficiformis* metagenome, a set of 1,366 spacers was detected, of which 1,349 were unique (Table 3). The largest CRISPR array in *P. ficiformis* contained 112 spacers. For the *S. foetidus* metagenome, a total of 723 spacers was identified, of which 714 were unique. Here, the longest array contained 67 spacers. Thirdly, in the *A. aerophoba* metagenome, a total of 9,669 spacer sequences were detected with 125 of these occurring more than once and with 9,547 being unique. The longest array found in *A. aerophoba* comprised 169 spacers. None of the spacer sequences were shared between the metagenomic samples suggesting that the three sponge microbiomes have their own distinct CRISPR systems.

With respect to potential targets of the spacers, the number of hits decreased from unknown targets, to plasmids, phages and to viruses in all samples (Table 3). Combining these results with the spacer taxonomy, most spacers originated in *Alphaproteobacteria* and *Actinobacteria*. All taxonomic groups of spacers had hits



**FIGURE 4 | Heatmap of defense mechanisms in (A) COG functional categories and (B) additional searches for the phage growth limitation and DNA phosphorothioation in the COG and Pfam databases.** The color scale ranges from 0 (black) to 2 (white) and indicates copies per megabase metagenome. Bray-Curtis dissimilarity is indicated by the dendrogram on top. Enriched feature relates to COG classes which are on average > 1.5-fold higher in seawater (circle) or in sponges (cross) over all sponge samples. Similar COG annotations are labeled on the left side of the heatmap and ordered from high to low copy numbers.

in the four target groups, with the largest amount of hits found for unknown targets. The three sponge metagenomes were shaped similarly with respect to spacer origins and targets (Figure 6). The proportion of spacer sequences originating from *Betaproteobacteria* was highest in *S. foetidus*, whereas *Gamma*- and *Deltaproteobacteria* were highest in *A. aerophoba*. *S. foetidus* showed more spacers originating from *Firmicutes* than the other sponge samples. Spacers from *Spirochaetes* were only found in the *A. aerophoba* metagenome. Overall, the distribution of spacers

with assigned taxonomy followed the genomic composition with correlation coefficients of 0.64, 0.59, and 0.88 (all  $p$ -values  $\leq 0.05$ ) for *P. ficiformis*, *S. foetidus*, and *A. aerophoba*, respectively (Figures 1 and 6). All spacers and direct repeat sequences are compiled in Supplementary Figures S2 and S3.

The number of unique direct repeats was 67, 40, and 218 for *P. ficiformis*, *S. foetidus*, and *A. aerophoba*, respectively (Table 3). In the datasets *A. aerophoba* and *S. foetidus*, three of the direct repeat sequences were shared, of which nine between



**TABLE 3 | Raw counts of identified CRISPR arrays and their taxonomic assignments, cas-genes, spacers and direct repeats.**

	Seawater	<i>Petrosia ficiformis</i>	<i>Sarcotragus foetidus</i>	<i>Aplysina aerophoba</i>
Pilerr-cr	5	90	76	384
CRT	9	108	83	529
Contigs with found Cas-genes	2	124	101	263
CRISPRFinder	0	77	47	290
CRISPR per megabase	0	0.21	0.25	0.62
CRISPR with assigned taxonomy	0	53	40	240
<i>Proteobacteria</i>	0	36	22	169
<i>Actinobacteria</i>	0	9	6	33
<i>Chloroflexi</i>	0	3	4	6
<i>Cyanobacteria</i>	0	1	1	6
<i>Firmicutes</i>	0	1	2	14
<i>Acidobacteria</i>	0	2	1	4
<i>Verrucomicrobia</i>	0	1	2	0
<i>Bacteroidetes</i>	0	0	1	7
<i>Deinococcus-Thermus</i>	0	0	1	3
CRISPR assigned to CAS-genes	0	26	20	144
Type I (A-F)	0	11	10	73
Type II (A-C)	0	1	2	10
Type III (A-B)	0	1	0	6
Type unknown	0	13	8	55
Largest array (# spacer)	0	112	67	169
Total number of spacer	0	1,366	723	9,669
Unique spacer	0	1,349	714	9,547
Spacer with found target	0	278	152	1,642
Phage	0	55	42	255
Virus	0	19	10	146
Plasmid	0	204	100	1,241
Target unknown	0	1,088	581	8,027
Total number of repeats	0	77	47	290
Number of unique repeats	0	67	40	218
Repeats with hits to CRISPRdb	0	55	25	144
CRISPRmap superclass A/B/C/D/E/F	0 – 0/0/0/0/0/0	47 – 1/7/19/2/6/2	21 – 0/6/8/0/5/2	88 – 3/30/36/0/23/7

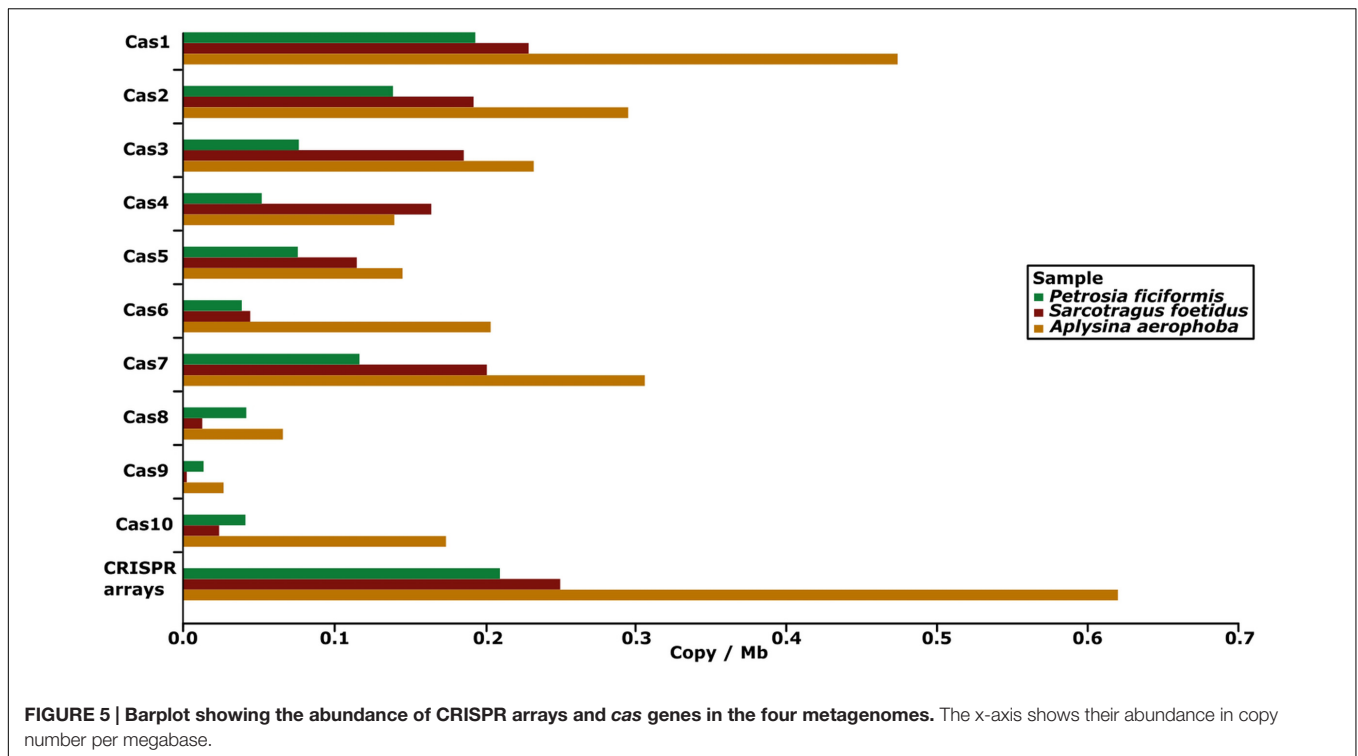
*P. ficiformis* and *A. aerophoba*, suggesting a horizontal transfer of either CRISPR arrays or bacteria. An amount of 55 (81.09%), 25 (62.5%), and 144 (66.06%) of *P. ficiformis*, *S. foetidus*, and *A. aerophoba* derived repeats were assigned to known repeat sequences using CRISPRdb. With respect to the classification using CRISPRmap, 47 (70.15%), 21 (52.5%), and 88 (41.12%) direct repeats for *P. ficiformis*, *S. foetidus*, and *A. aerophoba* were assigned to known superclasses, with the most abundant classes C, E, and B (Table 3). Notably, the superclasses were ordered decreasing in their conservation (Lange et al., 2013), showing a mixture of repeats with a roughly corresponding structure (superclasses B and C) and little sequence conservation (superclass E). Overall, 81.09 and 70.15% of all direct repeat sequences could be classified using CRISPRdb and CRISPRmap, respectively.

### Restriction Modification Systems

We identified a total of 3,057 RMSs in the metagenome datasets with 432 assigned to type I RMS, 2,379 to type II RMS, and 246 to type III RMS. A normalization of these raw counts to copies per megabase (cpm) resulted in a similar

distribution of RMS types I-III in the metagenomes. The sponge metagenomes showed higher abundances of all RMS types than seawater (2.48–5.08 cpm in sponges vs. 0.18 cpm in seawater). Type II was the most prevalent RMS type in the inspected metagenomes (*S. foetidus* = 4.03 cpm, *A. aerophoba* = 3.16 cpm, *P. ficiformis* = 2.06 cpm, seawater = 0.17).

The majority of type I RMS genes were assigned to *Proteobacteria*, *Actinobacteria*, and *Deinococcus-Thermus* in the sponge metagenomes, while in seawater, type I RMS was assigned exclusively to the classes *Beta-* and *Gammaproteobacteria* (Figure 7). The majority of type II RMS in sponge metagenomes was assigned to *Proteobacteria* (*Alpha-* and *Gamma-*) and *Actinobacteria* as well as to a lesser extent, to *Bacteroidetes*, *Cyanobacteria*, and *Acidobacteria*. The *S. foetidus* metagenome contained an unusually high number of type II RMS affiliated to *Actinobacteria*. Type III RMS was the most underrepresented group. Type III RMS in sponge metagenomes was most represented by *Alpha-* and *Gamma-Proteobacteria* as well as *Bacteroidetes* and *Chloroflexi*, while type III in the seawater sample was only represented by the *Alpha-* and *Gammaproteobacteria*, *Bacteroidetes* and the *Clostridia*.



## DISCUSSION

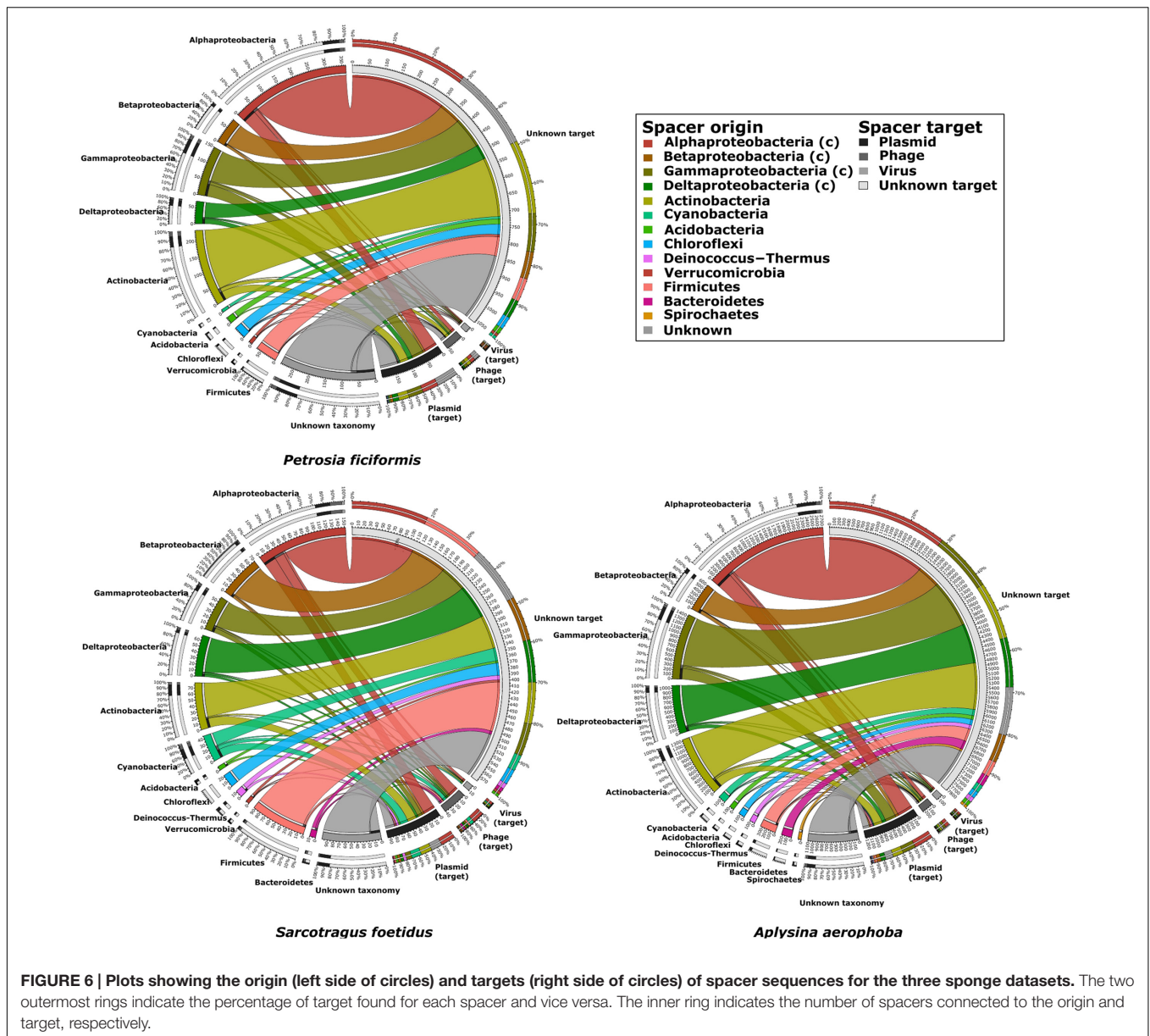
### General Features

We performed the taxonomic assignment of metagenomics reads by MG-RAST which has been previously attempted using microbial metatranscriptome data from a low microbial abundance sponge (Moitinho-Silva et al., 2014). While this approach offers the advantage of using the full metagenome dataset rather than a single gene marker (i.e., 16S rRNA gene), it may lose resolution for those phyla and candidate phyla where reference genomes are not available. Our results confirm previous findings that sponges harbor a distinct microbiota which is different from that of the surrounding seawater. Principal component analysis of the relative abundance of reads revealed a clustering of the sponge samples (Supplementary Figure S3). The sponge metagenomes overlapped in their composition and they showed a higher proportion of *Actinobacteria* and *Deltaproteobacteria* than seawater based on assignment of complete metagenomic reads. In contrast, the seawater metagenome revealed higher abundances of *Alphaproteobacteria*, *Flavobacteria*, and *Cyanobacteria* compared to the sponge metagenomes. With increasing availability of sequence data and the completion of draft genomes by single cell genomics (Kamke et al., 2014) or binning approaches (Gao et al., 2014; Burgsdorf et al., 2015), the assignment of complete reads rather than single gene markers should become widely acceptable.

The sponge metagenomes displayed much higher GC contents (58–63%) than the seawater metagenome (41%) (Figure 2A). As has previously been recognized, the prokaryotic GC content can be highly variable between different environments (Foerster

et al., 2005; Reichenberger et al., 2015), ranging from 34% for Sargasso Sea surface water samples to 61% for terrestrial soils. The GC composition of the sponge metagenomes is much higher than most other metagenomes, only to be superseded by metagenomes from saline ponds and contaminated soils (Reichenberger et al., 2015). While an explanation for the variation in GC composition remains wanting, there is increasing evidence that both, the phylogenetic composition of the samples and the environment shape the GC composition of the resident microbiota. With respect to the sponge metagenomes, the GC contents are likely a result of bacterial community composition. *Actinobacteria*, which are known for their high GC content, are much more prevalent in the sponge metagenomes than in seawater. Accordingly, *S. foetidus* displayed the largest abundance of *Actinobacteria* (Figure 1) and the highest GC content (Figure 2). Nonetheless, this cannot be the only explanation, because in spite of variable abundances of *Actinobacteria* within the three sponges (Figure 1), the GC content is very narrow (Figure 2A). Therefore, we posit that the specific microenvironment within sponges has some yet to be characterized effect on the microbial GC composition of sponges.

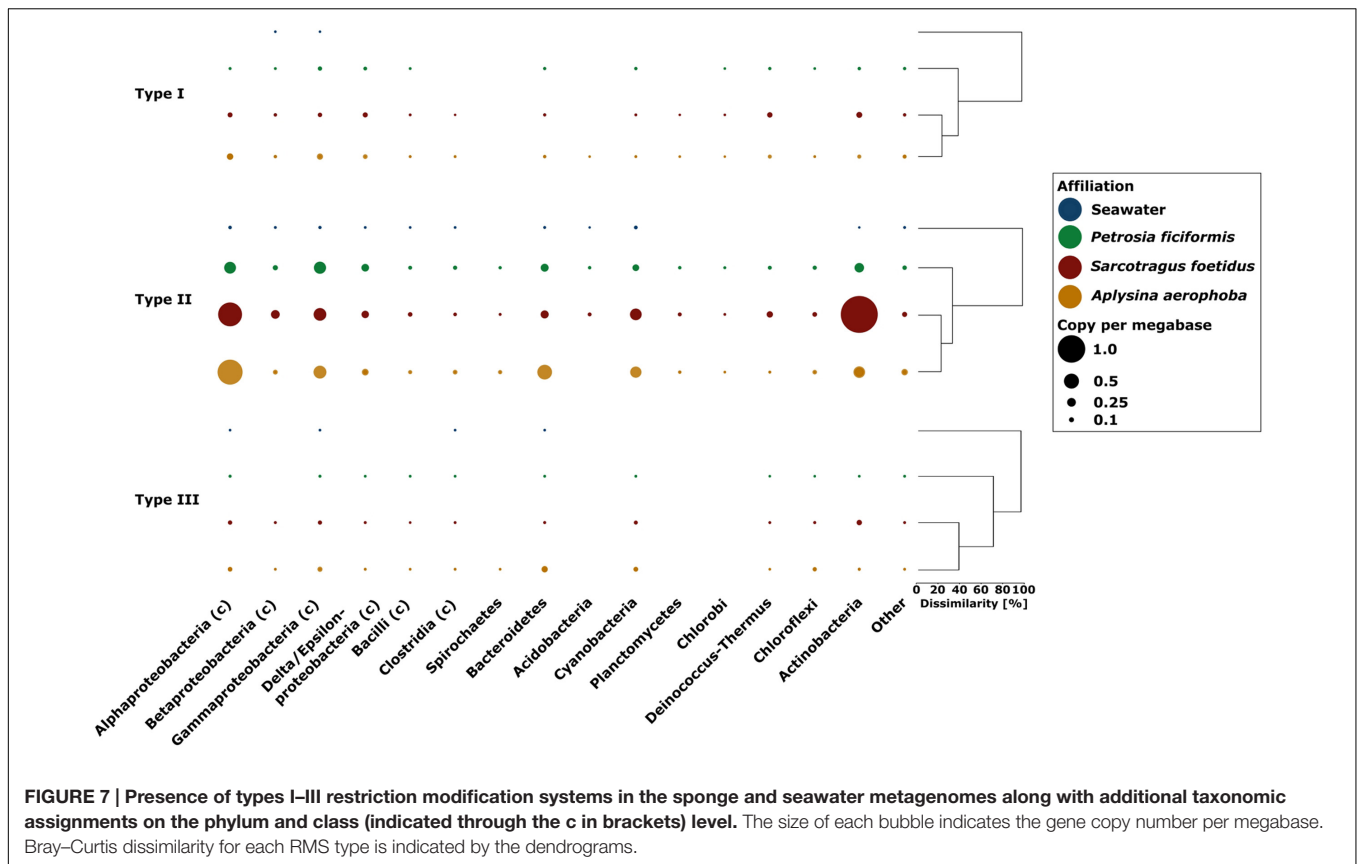
The sponge metagenomes displayed larger calculated average genome sizes (3.0–5.1 Mb) than that calculated for seawater (1.35 Mb) (Figure 2B; Table 2). The estimates for sponge bacterial genomes are on the larger end of genome size estimates derived from diverse metagenomic data (Giovannoni et al., 2014). It should however be noted that the comparison of closely related *Synechococcus* genomes from sponge symbionts versus those from seawater did not reflect this pattern (Burgsdorf et al., 2015). Larger genomes of sponge-associated bacteria may be the



evolutionary consequence of a more variable and nutrient-rich microenvironment within the sponge as opposed to the stable, nutrient poor seawater. Further the sponge-associated microbial consortia are constantly exposed to an ample source of free DNA resulting from the host's digestion of food bacteria. Whether and to what extent the mechanisms of horizontal gene transfer occur in sponges and whether this would then results in larger symbiont genomes remains to be investigated in future studies. The high prevalence of transposases and other mobile genetic elements within sponge microbiomes (Fan et al., 2012) does suggest that horizontal gene transfer is rampant in the sponge holobiont.

The overall functional annotation on the level of COG categories was more similar within the sponge samples than compared to the seawater sample (Figure 3). The functional

profile of *A. aerophoba* was more distant to the other sponge samples, which may have been influenced by a higher functional diversity as shown in the rarefaction curve (Supplementary Figure S2). Overall, only two COG categories were enriched in sponge metagenomes (defense mechanisms; cytoskeleton), while three COG categories were depleted in sponge metagenome over seawater (translation, ribosomal structure, biogenesis; cell motility; chromatin structure and dynamics). The category cytoskeleton was not pursued further owing to low gene abundance (<0.6 cpm). These results are somewhat different from previous data (Thomas et al., 2010), where the metagenome of the Australian sponge *C. concentrica* was enriched in two COG categories (secondary metabolites biosynthesis, transport and catabolism; replication, recombination, and repair) while being depleted in three other categories (translation, ribosomal



structure and biogenesis; nucleotide transport and metabolism; energy production and conversion in comparison to seawater). The only shared feature between these analyses is the depletion of sponge metagenomes in the category: translation, ribosomal structure, biogenesis. The category defense mechanisms is discussed in detail below.

## Defense Systems

The overall enrichment of the category defense mechanisms in sponge metagenomes over seawater metagenome is in agreement with earlier results, where functions related to viral defense were found to be enriched in sponges (CRISPR-Cas system, RMS) over surrounding bacterioplankton (Thomas et al., 2010; Fan et al., 2012) or in selected bacterial reference genomes (Burgsdorf et al., 2015). The defense system DNA phosphorothioation was, even though not functionally complete, also more prevalent in the sponge metagenomes. Further, genes associated with phage growth limitation (Pgl) were enriched in the sponge metagenomes. The microbial consortia within sponges may thus not only defend themselves against viruses and phages, but may also be capable of suppressing their growth. Since the Pgl system is only poorly characterized, further studies are needed to fully understand its potential impact on microbial communities.

Clustered regularly interspaced short palindromic repeats arrays were identified through a protocol using three different tools to avoid false positive hits. The total set of arrays (and direct repeats) was 77, 47, 283, and 0 for the metagenomes of

*P. ficiformis*, *S. foetidus*, *A. aerophoba* and seawater, respectively. While the CRISPR arrays in sponge metagenomes (0.21–0.62 cpm) are below the values described for completely sequenced genomes (0.72 cpm), they are still an order of magnitude above the values for seawater metagenomes, such as derived from the Sorcerer II Global Ocean Sampling expedition (0.042 cpm) (Sorokin et al., 2010). This value suggests a low number of CRISPRs in seawater and indeed, we found 0 hits in our seawater metagenome. The variation in the number of observed CRISPR arrays between the sponge metagenomic datasets may be due the fragmentation of generated contigs and is based on the used sequencing technology and the assembly algorithms (Gogleva et al., 2014).

The overall taxonomic assignment (> 68.83%) was comparable between the datasets with the largest fraction of CRISPR arrays affiliated to *Proteobacteria* followed by *Actinobacteria*, *Chloroflexi* and *Firmicutes* (Table 3). Similar results were observed for the origin of the spacer sequences (Figure 6). While this finding supports the presence of similar microbiomes within the different sponge species, only a small overlap of repeat sequences was identified. As ~30% of direct repeats could not be assigned to a superclass or known repeats, they may represent novel direct repeats. The fact that we did not detect any shared spacers suggests that the acquisition of protospacers may vary between bacterial individuals (Gogleva et al., 2014). The sponge-associated bacteria may either be exposed to different types of viruses, phages or plasmids (Burgsdorf et al., 2015) or to distinct viral

variations (Fan et al., 2012). With respect to the targets of the spacer sequences, their number decreased from the group “unknown” to plasmids, phages and viruses, and they were uniformly distributed among all identified phyla (Figure 6). The large fraction of hits to unknown and unique spacer sequences suggests that a large number of novel and diverse CRISPR targets and spacers can be expected in marine sponge metagenomes. The small overlap between spacers and direct repeats of the CRISPR-Cas systems likely reflects variations within each sponge metagenome as well as the specific acquisition of spacers from selected bacteria.

We found only ~50% of all CRISPR arrays adjacent to *cas* genes, which is likely an effect of the fragmentation of the assemblies. The *cas* genes were used to classify CRISPR systems into types and subtypes according to Makarova et al. (2011). Overall and as was expected, *cas1* (an universal marker of all CRISPR-*cas* systems) and *cas2* were most prevalent. CRISPR-Cas type I, described via *cas3*, was the most prevalent in all three sponge metagenomes followed by types II and III, identified by *cas9* and *cas10*. The latter two types were only found in very low abundances, suggesting type I to be the most important CRISPR type in the sponge microbiome. Interestingly, type I was also most prevalent in other environments such as the human gut (Gogleva et al., 2014) or groundwater (Burstein et al., 2016). The most abundant subtypes I–E showed a strong link to *Actinobacteria* (Makarova et al., 2015). In ecological terms, the high prevalence of CRISPR-Cas systems in sponge microbiomes may be necessary to defend the sponge-associated bacteria against viral particles that are drawn into the sponge holobiont by filtration. It has previously been estimated that the sponge-associated bacteria may be exposed to as many as 1000 viral particles per day (Thomas et al., 2010), thus an efficient defense against viral onslaught could be essential.

Restriction modification system have previously been shown to be more abundant in metagenomes from Australian sponges than in seawater (Fan et al., 2012). We here confirm these results for the Mediterranean sponges (2.48–5.08 cpm vs. 0.18 for RMS in seawater metagenome). The difference might be explained by the observation that larger genomes tend to have more RMS than smaller genomes (Makarova et al., 2013), which is indeed the case for the sponge metagenomes over the seawater metagenomes (Figure 2B). Among the different types of RMS, type II was most abundant in the metagenomes (Figure 7) which is consistent with previous findings for bacterial isolates (Oliveira et al., 2014). Similar to CRISPR, the RMS are mostly affiliated with *Alphaproteobacteria*, *Gammaproteobacteria*, *Betaproteobacteria*, and *Actinobacteria*. Both CRISPR and RMS thus appear as the first line of defense against foreign DNA, in particular against attack by viruses or phages.

## CONCLUSION

A comparison of microbial metagenomes from different Mediterranean sponge species versus seawater revealed bacterial defense systems as the consistently enriched feature in sponge metagenomes. These defenses include CRISPRs, RMSs, phage

growth inhibition and DNA phosphorothioation as the main mechanisms to combat foreign DNA from viruses, phages or other sources. The expanded genomic repertoire for bacterial defenses is likely the result of an evolutionarily long-standing adaptation where the resident sponge microbiota is exposed to free DNA resulting from the immense filtration activities of the animal host. In support of this, higher GC contents and larger calculated genome sizes were identified in sponge metagenomes over seawater. Collectively, our results indicate that the genomes of sponge microorganisms are/have been subject to horizontal gene transfer and that defense against foreign DNA is one prerequisite for an existence within sponges.

## AUTHOR CONTRIBUTIONS

HH designed the study, carried out metagenomics analysis, prepared figures and drafted the manuscript. BS collected the *A. aerophoba* sample and carried out its analysis. MJ participated in the analysis, preparation of figures and study design. KB sampled *P. ficiformis* and seawater and helped in the study design. LM-S sampled *S. foetidus* and participated in the study design. FF carried out analysis and wrote scripts. UA participated in the study coordination and drafting of the manuscript. UH coordinated the study and drafted/finalized the manuscript. All authors read and approved the final manuscript.

## FUNDING

This work was financially supported by the European Commission within its FP7 Program, under the thematic area KBBE.2012.3.2-01 with Grant Number 311932 (“SeaBioTech”) to UH and the DOE Joint Genome Institute with Grant Number CSP1291 to UH. BS and MJ were supported by a grant of the German Excellence Initiative to the Graduate School of Life Sciences, University of Würzburg.

## ACKNOWLEDGMENTS

We acknowledge the divers Panagiotis Vavilis and Thanos Dailianis (HCMR, Crete) for their diving support during the Milos sponge collection. We further thank the staff of the Marine Biology Station at Piran for their excellent support during our field work. We are grateful to Dr. Michael Richter (MPI Bremen) for help in first *P. ficiformis* metagenome assembly, and Dr. Rodrigo Costa (University of Algarve) as well as Dr. Alexander Keller (University of Würzburg) for helpful suggestions.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01751/full#supplementary-material>

## REFERENCES

- Abedon, S. T. (2012). Bacterial ‘immunity’ against bacteriophages. *Bacteriophage* 2, 50–54. doi: 10.4161/bact.18609
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comp. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Bayer, K., Schmitt, S., and Hentschel, U. (2008). Physiology, phylogeny and in situ evidence for bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*. *Environ. Microbiol.* 10, 2942–2955. doi: 10.1111/j.1462-2920.2008.01582.x
- Bell, J. J. (2008). The functional roles of marine sponges. *Estuar. Coast. Shelf Sci.* 79, 341–353. doi: 10.1016/j.ecss.2008.05.002
- Biswas, A., Gagnon, J. N., Brouns, S. J., Fineran, P. C., and Brown, C. M. (2013). CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA Biol.* 10, 817–827. doi: 10.4161/rna.24046
- Bland, C., Ramsey, T. L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N. C., et al. (2007). CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209. doi: 10.1186/1471-2105-8-209
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Burgsdorf, I., Slaby, B. M., Handley, K. M., Haber, M., Blom, J., Marshall, C. W., et al. (2015). Lifestyle evolution in cyanobacterial symbionts of sponges. *MBio* 6, e391–15. doi: 10.1128/mBio.00391-15
- Burstein, D., Sun, C. L., Brown, C. T., Sharon, I., Anantharaman, K., Probst, A. J., et al. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* 7:10613. doi: 10.1038/ncomms10613
- Edgar, R. C. (2007). PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* 8:18. doi: 10.1186/1471-2105-8-18
- Fan, L., Reynolds, D., Liu, M., Stark, M., Kjelleberg, S., Webster, N. S., et al. (2012). Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc. Natl. Acad. Sci. U.S.A.* 109, E1878–E1887. doi: 10.1073/pnas.1203287109
- Fieseler, L., Quaiser, A., Schleper, C., and Hentschel, U. (2006). Analysis of the first genome fragment from the marine sponge-associated, novel candidate phylum Poribacteria by environmental genomics. *Environ. Microbiol.* 8, 612–624. doi: 10.1111/j.1462-2920.2005.00937.x
- Fineran, P. C., Blower, T. R., Foulds, I. J., Humphreys, D. P., Lilley, K. S., and Salmond, G. P. (2009). The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc. Natl. Acad. Sci. U.S.A.* 106, 894–899. doi: 10.1073/pnas.0808832106
- Foerster, K. U., Von Mering, C., Hooper, S. D., and Bork, P. (2005). Environments shape the nucleotide composition of genomes. *EMBO Rep.* 6, 1208–1213. doi: 10.1038/sj.embor.7400538
- Gao, Z. M., Wang, Y., Tian, R. M., Wong, Y. H., Batang, Z. B., Al-Suwailim, A. M., et al. (2014). Symbiotic adaptation drives genome streamlining of the cyanobacterial sponge symbiont “*Candidatus Synechococcus spongiarum*”. *MBio* 5:e79–14. doi: 10.1128/mBio.00079-14
- Giovannoni, S. J., Thrash, J. C., and Temperton, B. (2014). Implications of streamlining theory for microbial ecology. *ISME J.* 8, 1553–1565. doi: 10.1038/ismej.2014.60
- Gogleva, A. A., Gelfand, M. S., and Artamonova, I. (2014). Comparative analysis of CRISPR cassettes from the human gut metagenomic contigs. *BMC Genomics* 15:202. doi: 10.1186/1471-2164-15-202
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007a). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8:172. doi: 10.1186/1471-2105-8-172
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007b). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360
- Hentschel, U., Piel, J., Degnan, S. M., and Taylor, M. W. (2012). Genomic insights into the marine sponge microbiome. *Nat. Rev. Microbiol.* 10, 641–654. doi: 10.1038/nrmicro2839
- Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167–170. doi: 10.1126/science.1179555
- Hoskisson, P. A., Sumbly, P., and Smith, M. C. M. (2015). The phage growth limitation system in *Streptomyces coelicolor* A(3)2 is a toxin/antitoxin system, comprising enzymes with DNA methyltransferase, protein kinase and ATPase activity. *Virology* 477, 100–109. doi: 10.1016/j.virol.2014.12.036
- Huson, D. H., Mitra, S., Ruscheweyh, H.-J., Weber, N., and Schuster, S. C. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* 21, 1552–1560. doi: 10.1101/gr.120618.111
- Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., Mcanulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Kamke, J., Rinke, C., Schwientek, P., Mavromatis, K., Ivanova, N., Sczyrba, A., et al. (2014). The candidate phylum Poribacteria by single-cell genomics: new insights into phylogeny, cell-compartmentation, eukaryote-like repeat proteins, and other genomic features. *PLoS ONE* 9:e87353. doi: 10.1371/journal.pone.0087353
- Lange, S. J., Alkhnbashi, O. S., Rose, D., Will, S., and Backofen, R. (2013). CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. *Nucleic Acids Res.* 41, 8034–8044. doi: 10.1093/nar/gkt606
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Li, C. W., Chen, J. Y., and Hua, T. E. (1998). Precambrian sponges with cellular structures. *Science* 279, 879–882. doi: 10.1126/science.279.5352.879
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477. doi: 10.1038/nrmicro2577
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569
- Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2013). Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.* 41, 4360–4377. doi: 10.1093/nar/gkt157
- Maldonado, M., Ribes, M., and Van Duyl, F. C. (2012). Nutrient fluxes through sponges: biology, budgets, and ecological implications. *Adv. Mar. Biol.* 62, 113–182. doi: 10.1016/B978-0-12-394283-8.00003-5
- Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2015). CDD: NCBF’s conserved domain database. *Nucleic Acids Res.* 43, D222–D226. doi: 10.1093/nar/gku1221
- Markowitz, V. M., Chen, I. M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2012). IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res.* 40, D115–D122. doi: 10.1093/nar/gkr1044
- Meyer, F., Paarmann, D., D’souza, M., Olson, R., Glass, E. M., Kubal, M., et al. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. doi: 10.1186/1471-2105-9-386
- Moitinho-Silva, L., Seridi, L., Ryu, T., Voolstra, C. R., Ravasi, T., and Hentschel, U. (2014). Revealing microbial functional activities in the Red Sea sponge *Stylissa carteri* by metatranscriptomics. *Environ. Microbiol.* 16, 3683–3698. doi: 10.1111/1462-2920.12533
- Nayfach, S., and Pollard, K. S. (2015). Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol.* 16:51. doi: 10.1186/s13059-015-0611-7
- Oliveira, P. H., Touchon, M., and Rocha, E. P. (2014). The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* 42, 10618–10631. doi: 10.1093/nar/gku734
- Ondov, B. D., Bergman, N. H., and Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12:385. doi: 10.1186/1471-2105-12-385

- Peng, Y., Leung, H. C., Yiu, S. M., and Chin, F. Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174
- Reichenberger, E. R., Rosen, G., Hershberg, U., and Hershberg, R. (2015). Prokaryotic nucleotide composition is shaped by both phylogeny and the environment. *Genome Biol. Evol.* 7, 1380–1389. doi: 10.1093/gbe/evv063
- Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., et al. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* 31, 1805–1812. doi: 10.1093/nar/gkg274
- Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2015). REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* 43, D298–D299. doi: 10.1093/nar/gku1046
- Ryu, T., Seridi, L., Moitinho-Silva, L., Oates, M., Liew, Y. J., Mavromatis, C., et al. (2016). Hologenome analysis of two marine sponges with different microbiomes. *BMC Genomics* 17:158. doi: 10.1186/s12864-016-2501-0
- Sorokin, V. A., Gelfand, M. S., and Artamonova, I. (2010). Evolutionary dynamics of clustered irregularly interspaced short palindromic repeat systems in the ocean metagenome. *Appl. Environ. Microbiol.* 76, 2136–2144. doi: 10.1128/AEM.01985-09
- Southwell, M. W., Weisz, J. B., Martens, C. S., and Lindquist, N. (2008). In situ fluxes of dissolved inorganic nitrogen from the sponge community on Conch Reef, Key Largo, Florida. *Limnol. Oceanogr.* 53, 986–996. doi: 10.4319/lo.2008.53.3.0986
- Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M. E. A., Mitros, T., et al. (2010). The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* 466, 720–726. doi: 10.1038/nature09201
- Taylor, M. W., Radax, R., Steger, D., and Wagner, M. (2007). Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol. Mol. Biol. Rev.* 71, 295–347. doi: 10.1128/MMBR.00040-06
- Thomas, T., Moitinho-Silva, L., Lurgi, M., Bjork, J. R., Easson, C., Astudillo-Garcia, C., et al. (2016). Diversity, structure and convergent evolution of the global sponge microbiome. *Nat. Commun.* 7, 11870. doi: 10.1038/ncomms11870
- Thomas, T., Rusch, D., Demaere, M. Z., Yung, P. Y., Lewis, M., Halpern, A., et al. (2010). Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME J.* 4, 1557–1567. doi: 10.1038/ismej.2010.74
- Vasu, K., and Nagaraja, V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.* 77, 53–72. doi: 10.1128/MMBR.00044-12
- Vogel, S. (1977). Current-induced flow through living sponges in nature. *Proc. Natl. Acad. Sci. U.S.A.* 74, 2069–2071. doi: 10.1073/pnas.74.5.2069
- Wang, L., Chen, S., Xu, T., Taghizadeh, K., Wishnok, J. S., Zhou, X., et al. (2007). Phosphorothioation of DNA in bacteria by dnd genes. *Nat. Chem. Biol.* 3, 709–710. doi: 10.1038/nchembio.2007.39
- Webster, N. S. (2007). Sponge disease: a global threat? *Environ. Microbiol.* 9, 1363–1375. doi: 10.1111/j.1462-2920.2007.01303.x
- Wehr, M., Steinert, M., and Hentschel, U. (2007). Bacterial uptake by the marine sponge *Aplysina aerophoba*. *Microb. Ecol.* 53, 355–365. doi: 10.1007/s00248-006-9090-4
- Weisz, J. B., Lindquist, N., and Martens, C. S. (2008). Do associated microbial abundances impact marine demosponge pumping rates and tissue densities? *Oecologia* 155, 367–376. doi: 10.1007/s00442-007-0910-0
- Wilson, M. C., Mori, T., Ruckert, C., Uria, A. R., Helf, M. J., Takada, K., et al. (2014). An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* 506, 58–62. doi: 10.1038/nature12959

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Horn, Slaby, Jahn, Bayer, Moitinho-Silva, Förster, Abdelmohsen and Hentschel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.